



113

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁶:

G06F 3/06, 13/40

A1

(11) International Publication Number:

WO 99/46671

(43) International Publication Date: 16 September 1999 (16.09.99)

(21) International Application Number: PCT/US99/05231

(22) International Filing Date: 10 March 1999 (10.03.99)

(30) Priority Data:
60/077,643 10 March 1998 (10.03.98) US(71) Applicant (for all designated States except US): QUAD RE-
SEARCH [US/US]; Suite 8, 9960 Indiana Avenue, River-
side, CA 92503-5457 (US).

(72) Inventor; and

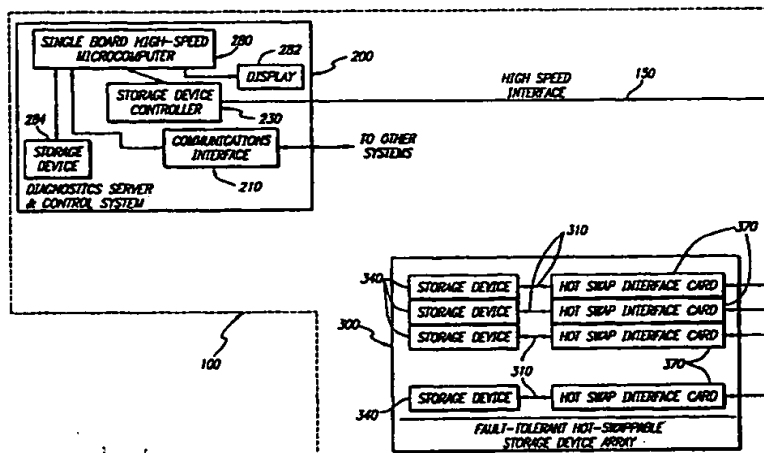
(75) Inventor/Applicant (for US only): DELLACONA, Richard
[US/US]; P.O. Box 8403, Newport Beach, CA 92658 (US).(74) Agents: PAUL, James, W. et al.; Fulwider Patton Lee & Utecht,
LLP, 10th floor, 10877 Wilshire Boulevard, Los Angeles,
CA 90024 (US).(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG,
BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB,
GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG,
KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK,
MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW,
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the
claims and to be republished in the event of the receipt of
amendments.

(54) Title: HIGH SPEED FAULT TOLERANT MASS STORAGE NETWORK INFORMATION SERVER



(57) Abstract

The information server system incorporates a high speed, microcomputer based server running industry standard operating system software enhanced to include functionality directed to operation of a new disk array controller, which controls the physically independent or integral disk storage device array, and communications interface. The disk array controller subsystem controls and communicates with the disk storage device array with a Fibre Channel protocol. The disk storage device array incorporates a plurality of disk storage devices with a corresponding number of bypass interface cards configured to facilitate the on-line addition, removal and replacement of disk storage devices. In addition to incorporating the above described buses and Fibre Channel capability, the disk array further incorporates a physically independent Fibre Channel compatible optical bus for high speed, communication between disk storage device array subsystem components, including the internal disk storage devices, independent from the information server. A wide array of user reconfigurable options are available as well as a scalable expansion capability.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

HIGH SPEED FAULT TOLERANT MASS STORAGE NETWORK INFORMATION SERVER

BACKGROUND OF THE INVENTION

Related Applications:

This application is based upon Provisional Patent Application Serial No. 60/077,643, filed March 10, 1998.

5

Field of the Invention:

This invention relates to a "plug and play", high speed, microcomputer based, Fibre Channel compatible and fault tolerant mass storage information server system. In particular, the present invention relates to a device and method for providing an enterprise-wide information server system which incorporates a dual loop arbitrated, Fibre Channel capable, multiple-fault tolerant, hot-swappable disk array requiring no host computer.

10

Description of the Prior Art:

15

Efforts have been made in the past to provide a mass storage file server capable of delivering information throughout an enterprise with high speed data throughput, scalable data storage capability in a convenient, easily configurable enclosure using well known, industry standard operating software. However, such systems have typically experienced many shortcomings and problems associated with the inability of presently available computer and communications hardware to sustain performance and survive failure of component devices. Such shortcomings have included the lack of capability to allow for scalability to accommodate increased storage requirements without the need to completely power down information server

20

25

system to facilitate installation of additional disk storage devices or communications capability. One of the solutions presently available requires that the entire information server system be taken off-line and powered down before any additional disk storage devices can be added to the disk storage array. This and other
5 comparable or more capable server systems require a significant amount of system administrator personnel time and server system resources to facilitate system maintenance for expansion of storage space, repair, and routine maintenance such as optimization and system health monitoring.

Other problems existing in the presently available information server
10 systems include a general inability of the present technology to provide continuous information server capability after some predetermined number of failures have occurred in the server system. Some prior art information server systems have provided a limited fault tolerance capability. Such systems typically employ a disk array server which incorporates extra disk space substantially in excess of that needed
15 by the enterprise serviced by the information server system. The extra disk space is incorporated into such systems with the addition of additional physical disk storage devices which are configured with a particular logical disk drive configuration tailored to meet the storage, fault tolerance and server requirements of the user.

To accomplish a desired level of fault tolerance, the disk array subsystem
20 is configured with various types of operating system software to create duplicate and multiple copies of the data stored on the information server system across various different combinations of the physical and logical disk drives. Any of a large array of fault tolerant disk array architectures are typically employed including, for example, various implementations of what is known the trade as the redundant array
25 of independent disks or "RAID" topologies and protocols. Upon detection of a complete or partial failure of a particular disk storage device, the operating system software program notifies the server of the failure, marks the portion of the disk which failed or logically removes the completely failed disk storage device from the disk array. Next, the operating system typically reallocates the remaining available
30 physical disk space into a modified configuration of logical disk drives.

The operating system then reconstructs new duplicate, multiple copies of the data stored on the server system within the constraints of the newly reduced amount of free disk space. The server system also alerts the server system administrator that a failure has occurred so that corrective action may be taken.

5 Depending on the exact nature of the failure, the necessary corrective action can include removal and replacement of the completely or partially failed disk storage device. The remove and replace system maintenance operation requires, in many systems, that the entire server system be taken off-line and powered down before the physical replacement operation can be performed. Some information server
10 systems permit removal and replacement of the defective disk storage device while the server system remains on-line and powered on. This process is commonly referred to by the trade as "hot-swapping" of devices. Such systems, however, require considerable hands-on intervention of the system administrator personnel to manually
15 manipulate the hardware interfaces and operating system software for purposes of physically and logically reintegrating the newly replaced disk device into the disk array subsystem. Also, the systems capable of hot-swapping, experience severe degradation of performance resulting from the process of taking corrective action.

 The process of logical reintegration requires a significant portion of the server system central processing unit and memory resources. These resources are
20 needed to accomplish the reallocation of the newly available free disk space into the logical disk configuration of the server and the redistribution of the multiple, duplicate copies of the information stored on the server across the new and remaining physical and logical disk drives. This need for server system resources, although temporary, results in a severe decrease in the performance of the information server system.

25 Thus, users of presently available information server technology have generally two types of information server system options available. The first type of server system is completely unavailable for the duration of the system maintenance operation. The second type of system is, in effect, unavailable to the users due to the seriously degraded performance experienced by the information server system during
30 the system maintenance operation. The following U.S. Patents, which are hereby

incorporated by reference in their entirety, appear to disclose various types and components of the above described information server systems: 5,402,428; 5,471,099; 5,479,653; 5,502,836; 5,517,632; 5,518,418; 5,522,031; 5,530,831; 5,544,339; 5,548,712; 5,615,352; 5,651,132; 5,659,677; 5,664,119; 5,666,337; 5,680,538; 5,694,581; and 5,701,406.

As a result of the problems and shortcomings of the technology incorporated into the presently available information server systems, users are left without a satisfactory server system which is capable of, among other features, continuous uninterrupted availability, nondegraded performance and simplified, quick and easy storage space expansion, reconfiguration, repair and routine maintenance. None of the previous devices have adequately met these needs. Thus, it is apparent that a need exists for an system which not only reduces or eliminates the shortcomings and problems associated with the currently available information server systems and related technology, but also which provides an efficient and cost-effective solution to such concerns.

SUMMARY OF THE INVENTION

The present invention is an information server system with a "plug and play", scalable, modular, fault tolerant, multi-loop, hot swappable architecture incorporating a central processing unit, a storage device controller connected to the central processing unit for controlling at least one storage device array and a communications interface system connected to the central processing unit for communicating with other systems. More particularly, the invention represents a plug and play storage system for information storage and retrieval applications and incorporates an on board computer server for the storage system, thus eliminating the requirement for resources from a host computer. The computer controls and communicates with a storage device controller and a communications interface with other systems external to the storage system. The storage device operates via a high speed interface to control an array of storage devices through their

individual hot swap interface cards.

In one presently preferred embodiment, the invention provides a self contained plug and play information server system which incorporates a high speed, microcomputer based, server running industry standard operating system software enhanced to include functionality directed to operation of an array controller for a storage device such as a magnetic disk array, optical device array, solid state memory or the like, and which controls the physically independent or integral storage device array, and a communications interface. In a presently preferred form, the array controller subsystem controls and communicates with the storage device array with a Fibre Channel protocol and topology compatible 1.0625 gigabit per second copper compact PCI and/or a fibre optic interface bus and an Intelligent Input/Output, "I₂O" bus for control of and communication with the disk storage device array.

The storage device array incorporates a plurality of storage devices with a corresponding number of bypass, or "bridging", interface cards configured to facilitate the on-line addition, removal and replacement of storage devices. In addition to incorporating the above described buses and Fibre Channel capability, the storage device array further incorporates a physically independent Fibre Channel compatible optical bus for high speed communication between storage device array subsystem components, including the internal storage devices, independent from the information server. The problems encountered with previously available information server systems are solved by the present invention, which can be manufactured relatively inexpensively from a variety of off-the-shelf hardware and software, either in standard configurations or on a custom configured basis. In either configuration, a wide array of user reconfigurable options are available as well as a scalable expansion capability.

The present invention accordingly provides for an information server system with a scalable, modular, fault tolerant, hot swappable architecture, that comprises a central processing unit; an array controller subsystem connected to the central processing unit for controlling at least one storage device array; and a communications interface subsystem connected to the central processing unit for

communicating with other subsystems of the information server and for controlling the array controller subsystem.

For convenience, and not by way of limitation, the invention will be described below in the context of magnetic disk storage devices as they represent readily available and compatible types of storage devices for information server applications. However, those skilled in the art will recognize that other storage devices and media such as optical disks, solid state memories or magnetic storage media would be applicable for various applications, depending on the state of development of the storage media technology and the application to which the system is to be put. Similarly, the invention will be described in the context of a Fibre Channel protocol and topology compatible 1.0625 Gigabit per second per second copper compact PCI and/or fibre optic interface bus and Intelligent Input/Output I₂O bus as the communications link between the storage device controller and the storage device array, although other communication links may be used, depending on the array architecture, communications speed requirements and available technology for data links.

In one presently preferred embodiment, the information server system further comprises a midplane connector for connecting interface cards for components. In a currently preferred aspect of the invention, each disk storage device array comprises a plurality of disk storage devices and a corresponding number of bypass interface cards, all of which communicate with one another and the information server. In another presently preferred aspect of the invention, each disk storage device array comprises a predetermined number of bypass interface cards which populate the entire information server system, whether or not the entire information server system is fully populated with a corresponding number of disk devices. Each disk storage device is preferably hot-swappable, and each disk storage device is mounted on a bypass interface card that connects to the midplane connector.

In a presently preferred embodiment, the disk array controller subsystem controls and communicates with one or more disk storage device arrays with an arbitrated dual channel Fibre Channel system, and each of the disk storage devices are

connected to the arbitrated dual channel Fibre Channel architecture, whereby each disk storage device may perform simultaneous reads and writes of data in response to any requests from the outside world through the information server. In another presently preferred aspect of the invention, the disk storage devices include electronic device registration devices, and the disk array controller subsystem monitors identification numbers of the electronic device registration devices. The disk array controller subsystem can thus monitor when a component in the information server system is removed or added. The electronic device registration devices are preferably integrated into an electronic circuitry of each of the disk devices such that engagement or disengagement of each disk storage device with the disk storage device array causes a triggering of the electronic device registration devices to generate and transmit a unique identifying serial number signal unique to each disk storage device to the disk array controller subsystem. A triggering of the electronic device registration devices preferably causes the electronic device registration devices to generate and transmit a unique identifying serial number signal unique to each disk device to the disk array controller subsystem, and upon receiving the signal, the disk array controller subsystem then immediately either initiates logical connection or disconnection of the disk device to or from the array, depending on whether the disk device has been engaged or disengaged, respectively. The disk array controller subsystem can thus accomplish the electrical and logical connection and disconnection of the disk device by control of the bypass interface cards. In another presently preferred aspect, the bypass interface cards comprise an independent but logically integrated optical bus for communication within the Fibre Channel topology and protocol between disk drives.

In another presently preferred embodiment, the disk array controller subsystem is adapted to configure one or more of the disk storage devices for a configuration selected from the group consisting of RAID 0, RAID 1, RAID 3, RAID 5, RAID 10 and an XOR RAID configuration.

In another presently preferred aspect of this invention, an operator may activate a so-called "hot button" which will disable the write function to the array,

thus preventing the writing of suspect data on the storage array after a system fault has been detected.

From the above, it may be seen that the present invention provides a novel, plug and play, high speed scalable and modular fault tolerant information server architecture which offers many benefits over prior art systems. Other features and advantages of the present invention will become apparent from the following detailed description of the invention, taken in conjunction with the accompanying drawings, which illustrate, by way of example, the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Referring now to the drawings wherein like reference numerals across the several different views refer to identical or corresponding parts,

FIG. 1 is a diagrammatic representation of the Information Server System of the present invention;

FIG. 2 is a diagrammatic representation of the Disk Array Controller subsystem of the Information Server System of FIG. 1; and

FIG. 3 is a diagrammatic representation of the Fault Tolerant, Fibre Channel Arbitrated Dual Loop, Hot-Swappable Disk Array subsystem of the Information Server System of FIG. 1.

DETAILED DESCRIPTION OF THE INVENTION

As can be understood from reference to FIGS. 1 - 3 in the accompanying drawings, which are provided for purposes of illustration and not by way of limitation, the information server system of the present invention, designated generally by reference numeral 100, incorporates two primary

subsystems, including an information server 200 and a disk storage device array 300 connected through its integral Fibre Channel architecture compatible, high speed physical and virtual backplane bus and Intelligent Input/Output, "I2O" bus to the integral and or independent and external bus 150. In the presently preferred embodiment, bus 150 is a Gigabit compact PCI/Fibre Channel interface and I₂O Bus, but other optical and copper based high speed data buses and the like are appropriate, depending upon system throughput requirements and available technology. The information server further incorporates secondary subsystems including, but not limited to, a communications interface subsystem 210, a disk array controller subsystem 230, and single board high speed computer 280 running a computer industry standard operating system software program such as, for example, Windows NT available from Microsoft Corporation.

The computer further communicates with a display 282, such as a touch screen display, and a storage device 284 such as a hard disk. The operating system software incorporates enhancements adding additional functionality directed to control of the communications interface 210, disk array controller 230 and single board computer 280. The single board computer ("SBC") 280 can include any of a wide number of suitable devices including, but not limited to, the Compact PCI CPU Board with Pentium Processor, Model No. ZT 5510, available from Ziatech Corporation. Modifications to enhance performance of the ZT 5510 can include an onboard 40 MB flash memory card for permanent storage of the non-reconfigurable portions of the Windows NT operating system software and an onboard, removable, PCMCIA 40 Mb flash memory card, "D2 FlashDisk" available from Sandisk Corporation for read/writeable storage of the reconfigurable portions of the Windows NT software. The respective manufacturer, Fibre Channel Consortium and I₂O Special Interest Group reference design data sheets and materials describing the detailed operating capabilities and specifications of each of these components are hereby incorporated by reference in their entirety.

The communications interface subsystem 210 communicates with

other systems on the above described buses and incorporates any combination of a number of widely available interface and communications tertiary subsystem circuit boards in the preferred embodiment. The interface boards incorporate, among other components, digital subscriber unit ("DSU") framers which can include, for example, T-1 communications compatible DSUs such as the Model No. DS2152 chip set from Dallas Semiconductor Corp. The communications boards incorporate, among other components, multi-channel communications processors which can include, for example, the 4-channel "Quick-Chip" communications processor from Motorola Corp. The communications circuit board also includes onboard buffer memory and a high speed, Fibre Channel compatible, optical bus for intra-server-component high speed data communications as well as Fibre Channel compatible, compact PCI bridge circuitry such as the bridge chip set available from Tundra. Additionally, for improved communications diagnostic, status and health monitoring capabilities, the communications board incorporates a bit error rate tester chip set, such as the device available as Model No. DS2172 from Dallas Semiconductor, configured for communication with the SBC 280 and operating system software, via the I2O bus. Although the communications interface subsystems 210 have been described with reference to well known T-1 communications architecture, the subsystems are adaptable for compatibility with nearly all other network and telecommunications architectures, protocols and topologies including, but not limited to, T-3, DS-3, OC-3C, OC-12C, OC-192C, FDDI, SONET, SCSI, TCP/IP, HiPPI and ATM. The respective manufacturer, consortium, industry association and organization reference design data sheets and materials describing the detailed operating capabilities and specifications of each of these components are hereby incorporated by reference in their entirety.

Referring generally now to the block diagram of FIG. 2, the disk array controller subsystem 230 further incorporates a high speed chip set, such as the ISP2100 Intelligent Fibre Channel Processor available from Qlogic Corp., which includes a reduced instruction set chip ("RISC") processor 235, a compact PCI

interface 240, and a Fibre Channel capable, arbitrated dual loop direct disk access protocol interface 250 for maximum communications bandwidth and redundancy. This controller 230 communicates data stored on the disk array or to be stored thereon through the communications interface subsystem 210 between the disk array 300 and the outside world. The controller 230 is fully Fibre Channel compliant and is capable of controlling up to 126 disk storage devices in the disk array 300. The controller 230 is adapted to configure the disk array 300 for fault tolerant, data redundancy protocols including, but not limited to, RAID 0, 1, 3, 5, 10 (1 & 0) and the new XOR RAID configuration. The respective manufacturer, consortium, industry association and organization reference design data sheets and materials describing the detailed operating capabilities and specifications of each of these components are hereby incorporated by reference in their entirety.

With general reference to FIG. 3, the fault tolerant, Fibre Channel capable, arbitrated dual loop, hot-swappable disk storage device array subsystem 300 further includes an enclosure having a front bay of slots for receipt of disk devices 340 and a rear bay for receipt of bypass / bridging interface cards 370. The midplane of the enclosure incorporates a backbone Fibre Channel compact PCI bus in addition to physically independent, but logically integrated (within the Fibre Channel topology and protocol) I2O and optical buses. The disk array 300 also incorporates a plurality of disk storage devices and a corresponding number of bypass interface cards 370, all of which communicate with one another and the information server 200 through any one or combination of the above described buses. Each disk array 300 preferably incorporates a predetermined number of bypass interface cards 370 which populate the entire enclosure of the array 300 whether or not the entire enclosure is fully populated with a corresponding disk devices 340. Thus, the bypass cards 370 maintain the logical and electrical integrity of the backbone buses of the array subsystem 300 regardless of the number of empty disk device 340 slots.

Each of the disk storage devices 340 may include any number of different storage media types including, but not limited to, optical, magnetic disk,

magnetic tape, magneto-optical media, flash and bubble memory devices, compact, read-only disk ("CD-ROM"), digital video disk and other similar technology mass storage devices. In one of the preferred embodiments of the present invention, each of the disk storage devices 340 includes a physical interlock system, not shown in the drawings, which is adapted to physically engage and disengage the disk storage device 340 to and from a physical enclosure which contains the disk array 300. Such devices can specifically include, for example, any one or combination of 4.3 Gb, 9.1 Gb and 18 Gb magnetic disk drives available from Seagate Technologies. Each of the disk devices 340 fully implement the arbitrated dual channel Fibre Channel architecture such that every disk device 340 may perform simultaneous reads and writes of data in response to any requests from the outside world through the information server 200. All of such devices of the present invention include, among other components, data carriers known to the trade as electronic device registration devices.

The data carrier device preferably incorporates a 64-bit ROM with a factory specified, unique 48-bit serial number. This capability enables unique identification of each and every physical disk device 340 manufactured for use with the present invention by any number of manufacturers, industry-wide for the foreseeable future. The data carrier device is integrated into the electronic circuitry of each of the disk devices 340 so that as the physical interlock of each disk device 340 is actuated to either engage or disengage the disk device 340 from the disk array 300, the interlock also triggers the data carrier device.

This triggering of the data carrier generates and transmits a unique identifying serial number signal unique to each disk device 340 which is transmitted, for example, over the I₂O bus to the disk controller 230. Upon receiving the signal, the controller 230 then immediately either initiates logical connection or disconnection of the disk device 340 to or from the array 300 depending on whether the disk device 340 has been engaged or disengaged, respectively. The controller accomplishes the electrical and logical connection and disconnection of the disk device 340 by control of the bypass interface card

370. Use of the data carrier device enables remote software controlled logical engagement or disengagement of each of the disk devices in addition to the above described operations. Thus, disk device 340 connection or disconnection can be accomplished either physically by a system administrator or remotely by error monitoring software functions of either the controller 230 or operating system running on the SBC 280. Such data carrier devices are readily available from a variety of manufacturers including the "iButton" available from Dallas Semiconductor as Part No. DS1990A.

The bypass/bridging interface cards 370 which populate the rear bay accept signals from the controller 230 to communicate via the Fibre Channel interface 150 and protocol between the storage devices, the diagnostics and the diagnostics and control system, and interfacing with the storage devices via the 310 connectors to the storage devices. Thus, each storage device can simultaneously perform multiple reads and writes within the arbitrated dual channel Fibre Channel topology and protocol capability previously described above. This capability allows remote software control to automatically detect insertion of new disk devices 340 into the disk array 300, and is thus hot swappable. Thus, automatic testing, formatting, and logical integration can occur without any intervention of administrator personnel, beyond physical insertion of a new bypass card 370 and/or disk device 340. Accordingly, when a new disk device 340 has been detected in the array 300 by receipt of the transmitted data carrier signal, unattended integration of the new disk into the array 300 topology is provided. Additionally, no degradation of the server system 100 performance will occur, since all such integration activity can be performed over the bypass card 370 optical bus. The circuitry of the bypass interface cards 370 typically incorporate any one of a number of Fibre Channel compatible bypass chip sets including, but not limited to, those available from Motorola Corp. as Part No. 10SX1189 and Vitesse Semiconductor Corp. as Part Nos. VSC7121 and VSC7122 Bypass Circuits.

For each of the disk storage device array 300 components, the respective manufacturer, consortium, industry association and organization reference design data sheets and materials describing the detailed operating capabilities and specifications of each of these components are hereby
5 incorporated by reference in their entirety.

The method of the invention embodies the control of hot swappable interface cards and associated storage devices arranged in an array and controlled via a high speed interface with a diagnostic and control system containing a communications interface and a storage device controller. A high speed computer
10 dedicated to system control and operation interfaces with both the communications interface and the storage device controller to sense and manage the condition of the various elements of the storage device array. By utilizing the method of the invention, the information server can be in continuous operation without the necessity of taking the system down to add storage capacity.

15 From the above, it will be appreciated that the present invention represents a new and novel arrangement for an information server system serving a computer information system. The system and method of the invention provide a plug and play, truly hot swappable storage device system that does not require an external host computer. The system of the invention can be directly connected
20 to Fibre Channel networks or Ethernet connections. The system provides both hot swappability and hot expandability for storage device addition and repair without powering down or additional diagnostic activities external to the system. Because of the scalability and modularity of the architecture, the system can be adapted to virtually any capacity.

25 While particular forms of the invention have been illustrated and described, it will also be apparent that various modifications can be made without departing from the spirit and scope of the invention. Accordingly, it is not intended that the invention be limited, except as by the appended claims.

WHAT IS CLAIMED IS:

1. An information server system with a scalable, modular, fault tolerant, multi loop, hot swappable architecture, comprising: a central processing unit including a digital processor, storage device and user interface; a storage array controller subsystem connected to the central processing unit for controlling at least one storage device array; and a communications interface subsystem connected to the central processing unit for communicating with other subsystems of the information server and for controlling said storage device array controller subsystem.
2. The information server system of Claim 1, further comprising a midplane connector for connecting interface cards for components.
3. The information server system of Claim 1, wherein each said storage device array comprises a plurality of disk storage devices and a corresponding number of bypass interface cards, all of which communicate with one another and the information server.
4. The information server system of Claim 1, wherein each said storage device array comprises a predetermined number of bypass interface cards which populate the entire information server system whether or not the entire information server system is fully populated with a corresponding number of storage devices.
5. The information server system of Claim 3, wherein each said disk storage device is hot-swappable.
6. The information server system of Claim 3, wherein each said disk storage device is mounted on a bypass interface card that connects to the midplane

connector.

5 7. The information server system of Claim 3, wherein said disk array controller subsystem controls and communicates with said at least one disk storage device array with an arbitrated dual channel Fibre Channel system.

10 8. The information server system of Claim 7, wherein each of said disk storage devices are connected to said arbitrated dual channel Fibre Channel architecture whereby each disk storage device may perform simultaneous reads and writes of data in response to any requests from the outside world through the information server.

15 9. The information server system of Claim 3, wherein said disk storage devices are selected from the group consisting of optical, magnetic disk, magnetic tape, magneto-optical media, flash and bubble memory devices, compact, read-only disk, and digital video disk storage devices.

20 10. The information server system of Claim 3, wherein said disk storage devices include electronic device registration devices.

11. The information server system of Claim 10, wherein said disk array controller subsystem monitors identification numbers of said electronic device registration devices.

25 12. The information server system of Claim 10, wherein said disk array controller subsystem monitors when a component in said information server system is removed or added.

30 13. The information server system of Claim 10, wherein said electronic device registration devices include a 64-bit ROM with a factory

specified, unique 48-bit serial number.

14. The information server system of Claim 10, wherein said electronic device registration devices are integrated into an electronic circuitry of each of the disk devices such that engagement or disengagement of each disk storage device with the disk storage device array causes a triggering of the electronic device registration devices to generate and transmit a unique identifying serial number signal unique to each disk storage device to said disk array controller subsystem.

15. The information server system of Claim 10, wherein a triggering of said electronic device registration devices causes said electronic device registration devices to generate and transmit a unique identifying serial number signal unique to each disk device to said disk array controller subsystem, and upon receiving the signal, said disk array controller subsystem then immediately either initiates logical connection or disconnection of the disk device to or from the array, depending on whether the disk device has been engaged or disengaged, respectively.

16. The information server system of Claim 10, wherein said disk array controller subsystem accomplishes the electrical and logical connection and disconnection of the disk device by control of said bypass interface cards.

17. The information server system of Claim 7, wherein said bypass interface cards comprise an independent but logically integrated optical bus for communication within the Fibre Channel topology and protocol between disk drives.

18. The information server system of Claim 10, wherein said disk array controller subsystem is adapted to configure said at least one disk storage

device for fault tolerant, data redundancy protocols.

5 19. The information server system of Claim 18, wherein said disk array controller subsystem is adapted to configure said at least one disk storage device for a configuration selected from the group consisting of RAID 0, RAID 1, RAID 3, RAID 5, RAID 10 and an XOR RAID configuration.

10 20. The information server system of Claim 1, wherein said communications interface subsystem comprises a digital subscriber unit framer.

21. The information server system of Claim 1, wherein said communications interface subsystem comprises an onboard buffer memory and a high speed, Fibre Channel compatible, optical bus for intra-server-component high speed data communications.

15 22. The information server system of Claim 1, wherein said communications interface subsystem comprises a bit error rate tester chip set.

20 23. The information server system of Claim 1, further comprising an operator controllable input device which disables the ability of the system to write new data on the storage devices of the storage device array.

25 24. The method of operation for an information server comprising: establishing high speed communication with a fault tolerant, hot swappable storage device array; sensing the status of each storage device of said array from information provided by a hot swap interface card associated with each storage device; analyzing information regarding the status of the storage device in a computer to assess operating characteristics of the storage device and overall operation of the information server; deriving signals for control of the storage devices to maintain system operation; communicating control signals to a service
30

device controller for control of individual storage devices; and transmitting the control signals over a high speed interface to individual hot swap interface cards to thereby control the operation of individual storage devices.

5 25. The method of Claim 24, further comprising the steps of: generation of a display of system status to a display panel; and receipt of inputs by the computer from a user interface.

10 26. The method of Claim 25, further comprising the step of: use of a touch screen display to both display status and accept inputs from the system user.

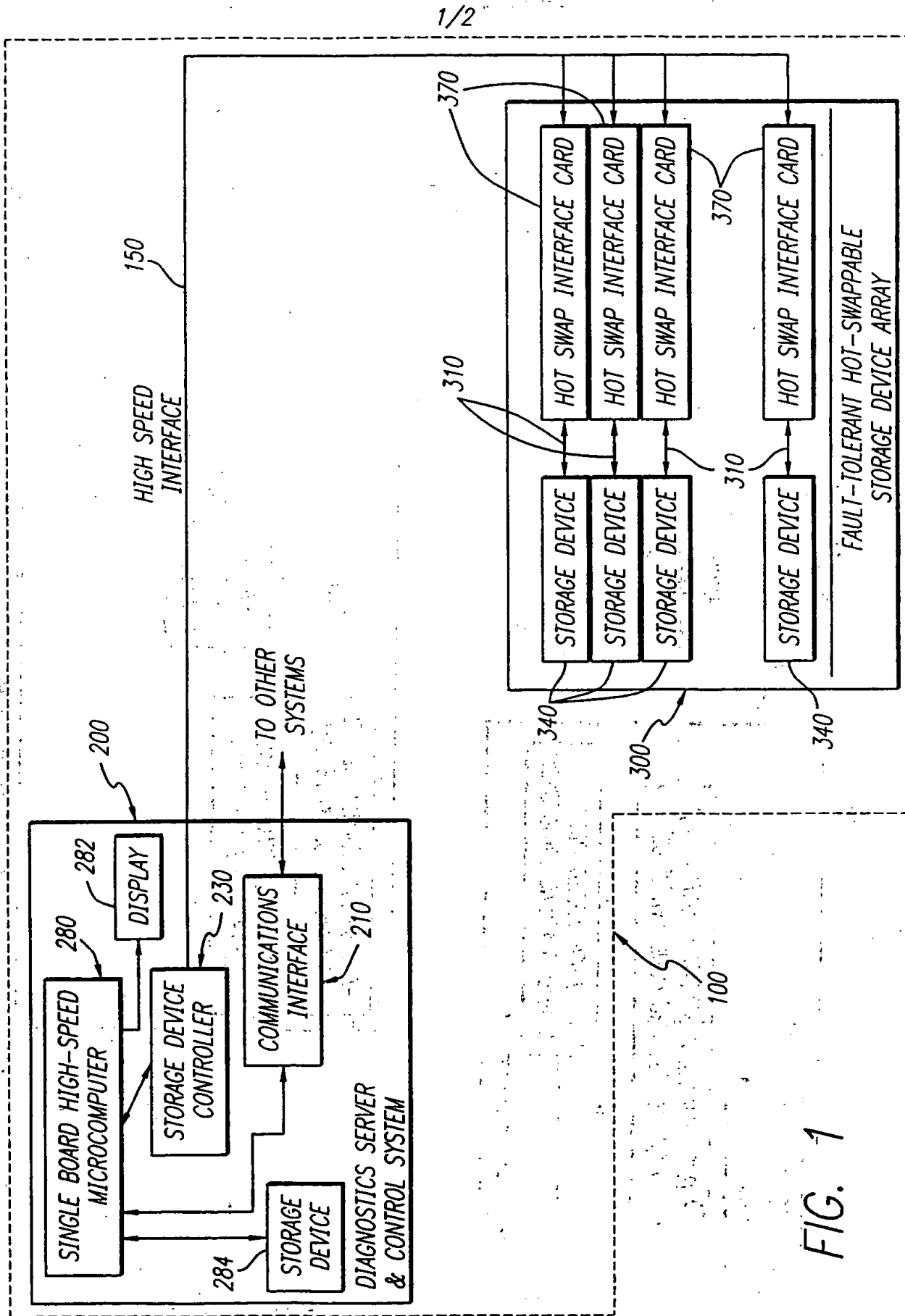
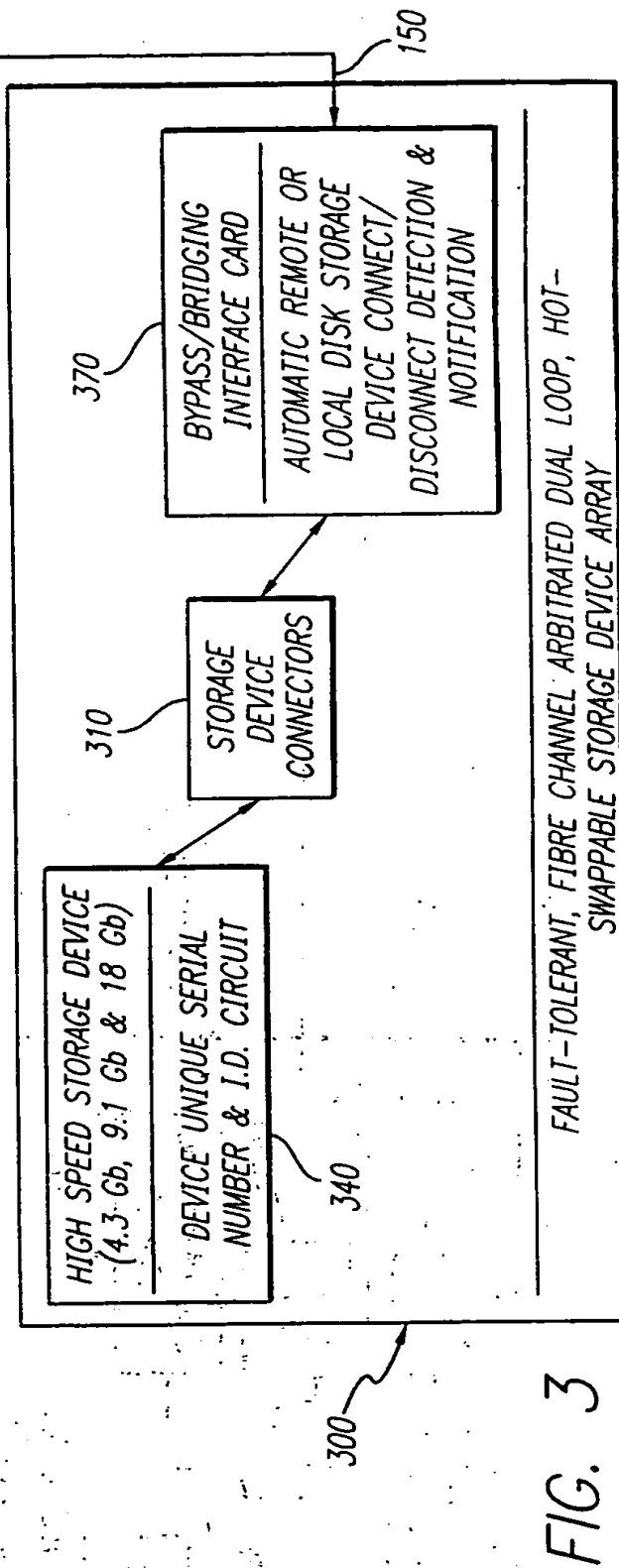
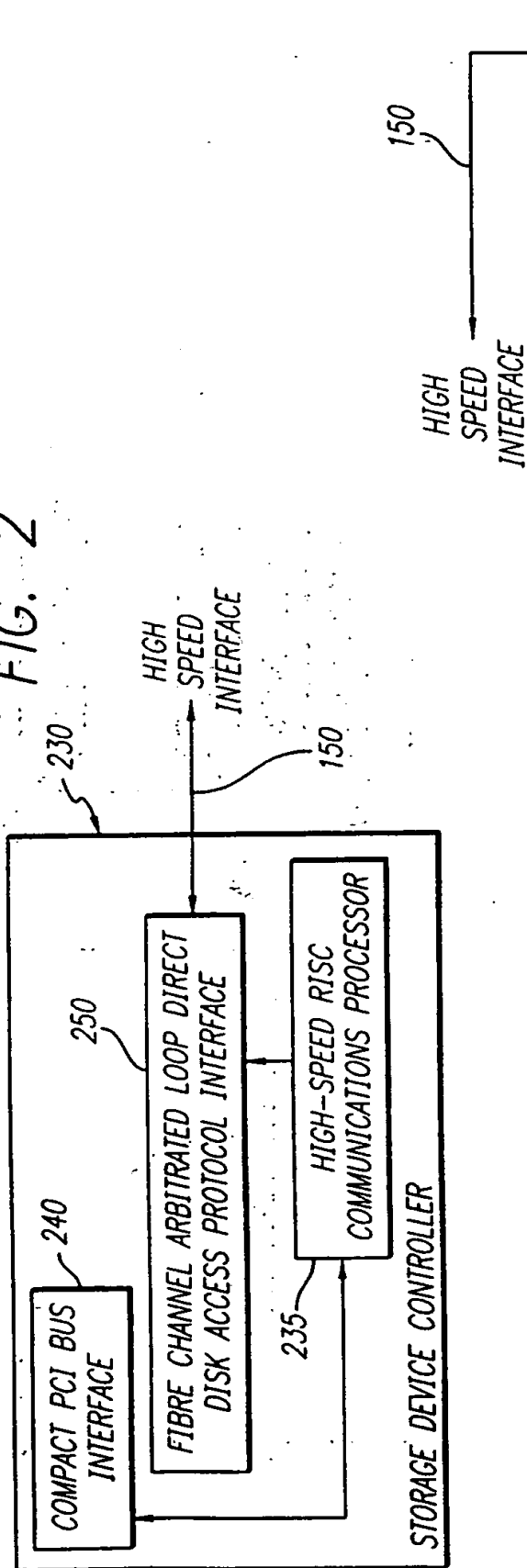


FIG. 2



INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/05231

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G06F3/06 G06F13/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 682 509 A (KABENJIAN) 28 October 1997 (1997-10-28) the whole document ---	1
A	WO 97 07458 A (EMC CORPORATION) 27 February 1997 (1997-02-27) the whole document ---	1, 24
A	R.W. KEMBEL: "IN-DEPTH FIBRE CHANNEL ARBITRATED LOOP" 1996, NORTHWEST LEARNING ASSOCIATES FOR SOLUTION TECHNOLOGY, USA (ISBN 0-0931836-81-6) XP002055479 page 269 - page 281 ---	1, 24
A	EP 0 768 599 A (COMPAQ COMPUTER CORPORATION) 16 April 1997 (1997-04-16) the whole document ---	1, 24
-/-		

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

30 July 1999

Date of mailing of the international search report

05/08/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3018

Authorized officer

Absalom, R

INTERNATIONAL SEARCH REPORT

Interr. .nal Application No

PCT/US 99/05231

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 479 653 A (JONES) 26 December 1995 (1995-12-26) the whole document	1,24
A	EP 0 541 996 A (FUJITSU LIMITED) 19 May 1993 (1993-05-19)	
A	EP 0 467 079 A (FUJITSU LIMITED) 22 January 1992 (1992-01-22)	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/05231

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5682509 A	28-10-1997	NONE	
WO 9707458 A	27-02-1997	US 5729763 A EP 0856172 A	17-03-1998 05-08-1998
EP 768599 A	16-04-1997	US 5809224 A	15-09-1998
US 5479653 A	26-12-1995	NONE	
EP 541996 A	19-05-1993	US 5369758 A DE 69224589 D DE 69224589 T JP 5210464 A US 5598528 A US 5751936 A	29-11-1994 09-04-1998 25-06-1998 20-08-1993 28-01-1997 12-05-1998
EP 467079 A	22-01-1992	JP 2056710 C JP 4049413 A JP 7086811 B AU 647132 B AU 7849091 A CA 2044869 A DE 69130092 D DE 69130092 T KR 9401667 B US 5721861 A	23-05-1996 18-02-1992 20-09-1995 17-03-1994 02-01-1992 20-12-1991 08-10-1998 21-01-1999 28-02-1994 24-02-1998